DOCUMENT RESUME

ED 121 814 95 TM 005 243

AUTHOR Bianchini, John C.; Vale, Carol A.
TITLE Investigation of the Appropriateness of the Anchor Test Study Equating Results for Selected Subgroups. Final Report.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Mar 75
CONTRACT OEC-0-74-8611
NOTE 25p.; For appendices, see TM 005 244-246; For related documents, see ED 092 601-634

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Bias; Caucasian Students; Elementary Education; *Elementary School Students; *Equated Scores; *Ethnic Groups; Grade 4; Grade 5; Grade 6; Negro Students; Reading; Reading Comprehension; *Reading Tests; Spanish Americans; Standardized Tests; Statistical Analysis; Test Interpretation; Vocabulary
IDENTIFIERS *Anchor Test Study; California Achievement Tests; Comprehensive Tests of Basic Skills; Gates MacGinitie Reading Test; Iowa Test of Basic Skills; Metropolitan Achievement Tests; Sequential Tests of Educational Progress; SRA Achievement Series; Stanford Achievement Tests

ABSTRACT

The Anchor Test Study (ATS) yielded equating tables for vocabulary, reading comprehension, and total reading scores for eight commonly used reading tests at the 4th, 5th, and 6th grade levels. Because of the original ATS sampling design which resulted in a nationally representative sample of school children at those grades, the equating tables might not be considered equally applicable for selected subgroups of the population--specifically, for black and Spanish-surnamed subgroups. The present study was done to determine whether use of the ATS equating tables for these two ethnic subgroups is warranted. Essentially, the analysis focused upon detecting interaction between test interrelationships and ethnic affilation in a way most relevant to the expected uses of the ATS equating tables--average differences between equivalent and obtained scores for each ethnic group, in various parts of the score range, were tested for significance. The detailed procedures for accomplishing the implied statistical tests are described. The analyses appear in Appendices A, B, and C (for grades, 4, 5, and 6, respectively) organized by vocabulary, reading comprehension, and total reading scores within grade level. The results did not indicate any systematic ethnic bias. The few isolated instances which did exist might be attributed largely to the sampling procedure used in the ATS; i.e., maximizing representation of the total population, rather than that of any specific subgroup. (RC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED121814

INVESTIGATION OF THE APPROPRIATENESS OF THE ANCHOR TEST STUDY EQUATING RESULTS FOR SELECTED SUBGROUPS

John C. Bianchini

Carol A. Vale

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM CS
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

FINAL REPORT

CONTRACT OEC-0-74-8611

March 1975

TM005 243

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

PROJECT REPORT

INVESTIGATION OF THE APPROPRIATENESS OF THE ANCHOR TEST STUDY EQUATING RESULTS FOR SELECTED SUBGROUPS

John C. Bianchini
Carol A. Vale

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
David Mathews, Secretary

Education Division
Virginia Y. Trotter, Assistant Secretary for Education

National Center for Education Statistics
Marie D. Eldridge, Administrator

FINAL REPORT

CONTRACT OEC-0-74-8611

March 1975

Educational Testing Service
Princeton, New Jersey
Berkeley, California

Prepared for the National Center for Education Statistics under contract number OEC-0-74- 8611 with the Office of Education, U.S. Department of Health, Education, and Welfare. Since contractors undertaking such projects are encouraged to express professional judgment, their points of view do not necessarily represent positions or policies of the Government, and no official endorsement should be inferred.

This Final Report is submitted by Educational Testing Service to DHEW/Office of Education under Contract OEC-0-74-8611 : Investigation of the Appropriateness of the Anchor Test Study Equating Results for Selected Subgroups. The ATS Equating and Norming Study Tapes prepared under this contract were sent under separate cover on February 25, 1975.

The original Anchor Test Study (Final Report, December, 1972) and the Supplemental Study carried out shortly thereafter (Final Report, November, 1973) yielded equating tables for vocabulary, reading comprehension, and total reading scores for eight commonly used reading tests at the 4th, 5th, and 6th grade levels. These tables, now published in the Anchor Test Study Users' Manual (U.S. Government Printing Office, 1974), have the same long-run applicability for any randomly selected child or random subgroup of that total population, by virtue of the original sampling design, which resulted in a nationally representative sample of school children at those grades. This statement remains true regardless of the ethnic membership of the individual child or the ethnic composition of the group. The tables may not be equally applicable, however, for selected subgroups of the population, in the sense that they may not be identical to those which would have resulted from an equating carried out on an appropriately selected sample from the subgroup population.

This study was designed to determine the applicability of the ATS equating results to selected ethnic subgroups of the total 4th, 5th, and 6th grade population--specifically, to black and Spanish-surnamed subgroups. Equating tables for ethnic subsamples comparable to those in the Users' Manual for the total sample do not exist and cannot be determined because the ATS sampling design does not provide ethnic subsamples suitable for

this purpose. Specifically, the ethnic subsamples are comparatively small and highly concentrated in a few major strata, hence from a relatively small number of schools. By the random test assignment process, the samples of black and Spanish-surnamed children were unequally allocated to the 22 different test-pairs involved in the study and unevenly dispersed with respect to ability level. The resulting ethnic samples for any given test-pair can thus be expected to be substantially less reliable than are the samples for white students. Because of these distributional problems, the question of whether different equating tables might be generated for different ethnic subgroups cannot be addressed by the ATS data.

To determine, then, whether use of the ATS equating tables for all ethnic groups is warranted, the following approach was adopted in this study.

Consider any pair of tests, X and Y. For a given score value of X, designate Y' as the equivalent score on test Y yielded by the ATS equating tables. The Anchor Test Study provides independent random samples of whites, blacks, and Spanish-surnamed, for each of which, scores on tests X and Y are available. For each score value of X, then, there is an equivalent score Y' and a distribution of observed scores on test Y, for each of the ethnic subgroups. The question of equal applicability of the equating tables then may be rephrased to ask whether the differences between the observed Y scores and the equivalent Y' scores are equal for all ethnic subgroups.

Asking the question in this way introduces sources of variance other than that due strictly to the equating process; for example, differences between racial groups may be exaggerated or attenuated by differences in measurement error. It does not, moreover, address the question of the "validity" of Y' for any subgroup or for the total population. But the

critical aspect of the question is one of ethnic "bias:" Given that a child has a particular score on test X, is his expected score on test Y dependent upon his ethnic membership. If not, then Y' may be considered equally applicable to ethnic subgroups. If there are significant differences by race in the expected Y, given X, then Y' is not equally applicable to all ethnic subgroups.

It is probably worth noting that this procedure does not require independence of the equating samples and the ethnic subsamples used to test the condition of equal applicability of the equating tables to all ethnic subgroups. That is, we need not have random "holdout" samples of whites, blacks, and Spanish-surnamed independent of those upon which the equating tables were obtained, since the procedure is essentially a test of the equality of the expected Y values, for the three subgroups.

Essentially, then, this phrasing of the question focuses upon detecting interaction between test interrelationships and ethnic affiliation in a way most relevant to the expected use of the ATS equating tables. If the purpose of the tables is to provide the user with an equivalent Y' in lieu of an actual Y, then the approach used here will determine whether this substitution is subject to ethnic bias. The detailed procedure for accomplishing the implied statistical tests is described below.

Procedure

The analysis to be described was carried out separately by grade, for each subtest and total test score, and for each test-pair in both directions, for a total of 396 applications (3 grades x 3 subtests x 22 test-pairs x 2 directions). For a given test-pair, the two orders of administration were

combined. Because the Gates-MacGinitie was actually paired only with the MAT for administration in the ATS Supplemental Study, only this equating was subjected to analysis. For all seven tests originally included in the Anchor Test Study, the equatings to each of the other tests were analyzed.

Information on the ethnic membership of each pupil was obtained from the response to question 2 of the "For Test Administrator Use Only" portion of the answer sheet. Nonrespondents were excluded from the analyses.

With pupils grouped into white, black, and Spanish-surnamed for each analysis, differences among the discrepancies* between the observed score on test Y and the equivalent score Y' for each subgroup were tested for significance as follows.

First, an overall conventional one-way analysis of variance for differences among the three groups was carried out, routinely followed by post-hoc comparisons (Dunnett's t) of each minority group vs whites. It is known, however, that the size of equating error is a function of score level, with error being generally larger at the lower (chance) portion of the score range (ATS Final Report, Tables 5-15, 5-16, 5-17 of the Project Report). This, combined with the distributional problems (discussed earlier) associated with the black and Spanish-surnamed samples for some test-pairs, suggested that an overall test which does not take score level into account might generate significant differences solely as a result of disproportionate sample representation at lower score levels.

---

*Throughout this discussion, the term discrepancy refers to the difference between equivalent scores and actual scores on test Y. The term difference is used in the conventional way to refer to differences between ethnic group means--in this case, the average of the discrepancies between equivalent scores and actual scores on test Y for the three ethnic groups.

Therefore, each test-pair was also subjected to an (unweighted means) two-way analysis of variance (race by score sub-range on the Y scale to identify specific equivalent score ranges yielding different results for the three subgroups.) To avoid extremely small samples at most score levels, each score range was divided into fifths; these sub-ranges were then used as a factor in the two-way analysis. Even with these collapsed score ranges, however, sample size was sometimes too small to include either the top or bottom levels in the analysis, and in approximately half of the analyses, the Spanish-surnamed group had to be deleted entirely. In such cases, a reduced two-way analysis was carried out for the remaining two groups. Post-hoc comparisons of blacks vs whites and Spanish-surnamed vs whites at each selected score level of the Y scale were carried out, again using the Dunnett procedure. This sort of analysis permits statements about differences, if they exist, in any of the selected sub-ranges, and about systematicity of differences across all or portions of the score range, as well as determining whether the tables are biased with respect to blacks, or Spanish-surnamed, or both (as compared with whites).

The decision to collapse into fifths of the score range for all tests was, of course, arbitrary and cannot be expected to be optimal for all (or even any) test-pairs. But, to the extent that the within-sub-range distributions are similar among the groups and/or the $Y' - \overline{Y}$ discrepancies are uniform, collapsing presents no special interpretational difficulty. If these conditions do not hold, collapsing within these relatively small score ranges still permits pinpointing the approximate location of the large discrepancies.

An unweighted means analysis of variance, rather than the alternative least squares procedure, was used to keep distributional imbalances from unduly influencing the estimates of effects attributable to score level, race, and interaction, and thus the concomitant significance tests.

For each of the types of analysis of variance, and for all follow-up tests as well, cell means and variances were computed using the original ATS individual sampling weights to obtain best population estimates for these statistics. For the testing procedure itself, each sample was regarded as a random selection from the respective population and unweighted N's were used.

Results and Discussion

The detailed results of the analyses appear in Appendices A, B, and C (for grades 4, 5, and 6, respectively) of this report, organized by vocabulary, reading comprehension, and total reading scores within grade level. For each analysis, a two-page set of results is presented. The test-pair shown in the heading is so ordered as to indicate the direction of the equivalent score transformation; i.e., actual scores on the first member of the pair, X, were transformed to equivalent scores on the second member, Y. The first page indicates, for each of the three ethnic groups, for each of the five score sub-ranges and for the total score range, the set of descriptive statistics, including the actual N, the weighted N, the mean discrepancy, and the standard deviation.

The second page contains the summary of individual comparisons of blacks vs whites (top) and Spanish-surnamed vs whites (bottom) at each score level included in the two-way analysis. Those instances for which the entire Spanish-surnamed group was deleted are so indicated. The highest

score level was deleted in about 75% of the two-way analyses, and the Spanish-surnamed group, in about half. It was only occasionally necessary to delete the lowest score level. The post-hoc comparisons for the one-way analyses appear below those for each score level, under the heading "total score range."

The results of the two-way analyses are summarized in Tables 1-9, for vocabulary, reading comprehension, and total reading score for grades 4, 5, and 6. These tables contain all the statistically significant* results for score sub-ranges, designated as level 1 (lowest fifth of the score range) to 5 (highest fifth of the score range). Whenever discrepancies are positive for one subgroup and negative for the other, differences between subgroups are exaggerated. In such cases, only subgroup differences which are significant in terms of the absolute value of the average discrepancy are included in these tables. Entries above the diagonal in the upper left corner of a given cell present significant differences for the black (B) vs white (W) analyses, and those below the diagonal in the lower right portion of the cell, for the Spanish-surnamed (S) vs white comparisons. For each score level, the group with the greater discrepancy in absolute value is indicated.

Some conclusions may be generalized over all three grades and all test pairs. First, the score level differences were significant in virtually every two-way analysis, with the greatest differences for all ethnic groups in the lowest (chance) score level. While this is not, in itself, of particular interest to this study, it does reinforce the original ATS findings

---

* The 1 percent level of significance was adopted as the criterion for determining statistically significant differences for the F tests in the analyses of variance and for the Dunnett t-tests in the post-hoc comparisons.

with regard to equating error and supports the decision to incorporate score level as a factor in the present study to aid the interpretation of significant differences in the overall score range analyses.

Second, race and/or interaction effects were significant in about 17% of the overall analyses. The post-hoc tests which followed up those analyses involved some 3800 ethnic group by score level comparisons, and showed approximately as many statistically significant instances of greater discrepancy for whites as for blacks (162 and 168, respectively), while the Spanish-surnamed subgroups had larger discrepancies in only 28 cases. It should be noted that these 338 statistically significant differences between ethnic subgroups at designated score levels represent less than 10% of the total number of comparisons involved in these post-hoc analyses.

Where significant differences occur, in the 4th and 5th score levels they almost always indicate a larger discrepancy for blacks or Spanish-surnamed, ranging from about 3 to 6 raw score points. However, such differences are infrequent and always based on extremely small N's (usually under 20 students). At score levels 1 and 2, usually the white sample has the larger discrepancy, generally 5 to 8 raw score points at level 1 (chance) and about 2 to 3 points at level 2. Significant differences at score level 3 generally indicate that the minority groups have the largest discrepancies (about 2 to 5 raw score points). In this middle portion of the score range, sample size is usually sufficiently large for the estimates to be relatively stable, but they are too sporadic to be meaningful. Furthermore, significant differences from the two-way analyses are not consistent across score levels for any given analysis,

across the several analyses for any given test, or across grade levels.

Of the 396 conventional one-way analyses of variance, only 10% yielded statistically significant results. A detailed review of all post-hoc comparisons between whites and each minority group clearly shows that in about 95% of the cases, not only are the comparisons statistically insignificant, but also the average discrepancy between equivalent and actual scores for all tests is quite small for each group, generally only a fraction of a raw score point.

All the statistically significant post-hoc comparisons for the one-way analyses are summarized in Table 10. Only test-pairs for which post-hoc results are significant are shown in the table. Those pairs for which significant differences occurred in both directions are placed together. The body of the table contains the group size and the magnitude and direction of the average discrepancy between equivalent and actual scores over the entire score range for the subgroup with the larger value, identified by the letter B for the black subgroup and S for the Spanish-surnamed subgroup.

It is clear from Table 10 that even where the between-group differences are significant, the size of the larger discrepancy is still small, the largest being 3.85. Considering that the minority subgroup sizes are generally small, especially for cluster samples, differences of 1 to 2 raw score points would be regarded as having no significant educational meaning. In the few isolated instances where the differences are greater than two raw score points, the samples are even smaller; all but four are less than 125 students.

As with the results of the two-way analyses, these results do not indicate any systematic ethnic bias. The few isolated differences which do exist may be attributed largely to the sampling procedure used in the ATS; i.e., maximizing representation of the total population, rather than that of any specific subgroup.

Thus, the ATS equivalent score tables are judged to be equally appropriate for all three ethnic groups. This, however, does not imply that each test is equally appropriate for any group of children, and that test users can ignore differences among tests. Indeed, users of the data provided in the ATS Users' Manual are urged to select the most appropriate test on the basis of educationally relevant group characteristics, reading curriculum, and test characteristics, such as content, skills measures, and level of difficulty. In so doing, the users will maximize the appropriateness of both the initial measurement and any subsequent score transformation resulting from the use of the Anchor Test Study equivalent scores.

TABLE 1

Significant Post-Hoc Comparisons for Score Levels from Two-Way Analyses*

Grade 4: Vocabulary

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT				4 B	3 B 2 W		
CTBS						4 B 3 B	2 W
ITBS		4 B 3 B 2 W		4B			3 B 3 S
MAT	2 W	5 B 2 W	2 W 4 S			2 B 1 W	2 B
STEP (II)		2 W		3 B			2 W 1 W
SRA	3 B		3 B	4 B 3 B	3 B 2 W 1 W 1 W		4 B 3 B 1 W
SAT			2 W		3 B	4 B 3 B	

GMT \ MAT	MAT
GMT	4 B 2 W

MAT \ GMT	GMT
MAT	2 W 1 W

* Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 2
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 4: Reading Comprehension

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT		3 B		2 W 2 W	1 W	3 B	
CTBS	1 W				3 B	3 B 2 W	4 B
ITBS		4 B 3 B		3 B			3 B 2 W
MAT	3 B 2 W		1 W		3 B 2 W	2 B 1 W	
STEP (II)	2 W			4 B			
SRA	4 B	4 B 3 B 2 W	2 W				
SAT	4 S 3 S 2 W	3 B 2 W	2 W 1 W		3 B 1 W	3 B	

	MAT	GMT
GMT		
MAT		2 W 1 W

*Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 3
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 4: Total Reading

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT		3 B 2 W			2 W		
CTBS							
ITBS		4 B 3 B		4 B 3 B 2 W		3 S	3 B
MAT	2 W	4 B 2 W	4 S		2 W	2 B 1 W	
STEP (II)				4 B			
SRA	4 B	4 B 3 B 2 W		4 B	2 W		
SAT					4 B 3 B 2 W	4 B 3 B	

*Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 4
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 5: Vocabulary

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT		2 W		2 W	4 B 2 W	2 W	3 B
CTBS	5 B 4 B 3 W		1 W			4 B 3 B	3 B 3 S
ITBS	4 B 2 W	3 B 1 W		3 B 1 W		1 W	1 W 1 W
MAT	2 W	4 B 2 W			4 B 3 W 2 W	3 B	1 W
STEP (II)		4 B				2 W 2 W	3 B
SRA	5 B 2 W	4 B 2 W		2 W	2 W		3 B 1 W
SAT	5 B	2 W		1 W			

	MAT
GMT	2 W

	GMT
MAT	4 B 2 W 4 S 2 W

*Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed—the group with the higher absolute value of the discrepancy.

TABLE 5
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 5: Reading Comprehension

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT		4 B		2 W	2 W	2 W	3 B 2 W
CTBS			3 S		2 W	3 B	
ITBS		4 B 3 B 2 W / 2 W		2 W			
MAT	3 B	4 B / 4 S	3 B			3 B 2 W	1 S
STEP (II)	2 W /	4 B					
SRA	3 B 2 W /	4 B 3 B 2 W	3 B	2 W	2 W / 4 S		
SAT	3 B / 4 S			4 S	4 B 3 B 2 W	3 B 2 W	

	MAT	GMT
CMT	4 B 2 W	
MAT		3 B 2 W

*Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 6
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 5: Total Reading

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT		2 W			4 B 2 W	2 W	
CTBS						4 B	
ITBS		4 B 3 B 2 W 4 S		2 W			
MAT		4 B				3 B 2 W	2 S
STEP (II)		4 B					
SRA	2 W	2 W			4 S		
SAT		4 B 2 W	3 B 2 W		4 B 3 B	4 B	

	MAT	GMT
GMT	2 W	
MAT		4 B

*Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 7
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 6: Vocabulary

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT			3 B		4 B 2 W	1 W	2 W
CTBS	3 B		2 W		4 B 2 W	3 B 2 W	
ITBS							
MAT	2 W		3 S		3 W		
STEP (II)	2 W	4 B 3 B				2 W 2 W	4 B 2 W
SRA	4 B 2 W	3 B	2 W 3 S		5 B 4 B 5 S 4 S 3 W		2 W
SAT	3 S	4 B	4 B 3 B 2 W 1 W		4 B	4 B	

	MAT
GMT	

	GMT
MAT	4 B 2 W

*Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 8
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 6: Reading Comprehension

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT		4 B				2 W	4 S
CTBS	3 B 2 W			3 B 2 W	4 B	4 B 2 W	
ITBS	3 B 2 W	4 B 3 B 2 W		3 B 2 W			3 B 2 W
MAT	3 B 2 W 1 W					4 B 3 B	
STEP (II)	4 B 2 W	3 B				3 B 2 W 4 S	2 W
SRA	2 W	3 B	1 W 1 W	2 W	5 B 4 B 3 B 2 W 3 S 2 W		2 W
SAT	3 S		2 W	2 W	4 B	3 B 2 W	

	MAT	GMT
GMT		
MAT		4 B 2 W

*Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 9
Significant Post-Hoc Comparisons for Score Levels
from Two-Way Analyses*

Grade 6: Total Reading

	CAT	CTBS	ITBS	MAT	STEP (II)	SRA	SAT
CAT		3 B			3 B	4 B	
CTBS	3 B			3 B 2 W	4 B	2 W	
ITBS				3 B		2 W	3 B
MAT	3 B 2 W		2 W			3 B	
STEP (II)	4 B 2 W					2 W	3 B
SRA	2 W		2 W		4 B 4 S		
SAT			2 W	2 W 4 S	4 B		

	MAT	GMT
GMT		
MAT		4 B 2 W

* Levels range from 1 (lowest fifth of score range) to 5 (highest fifth). B, W, or S indicate, respectively, blacks, whites, or Spanish-surnamed--the group with the higher absolute value of the discrepancy.

TABLE 10

Average Discrepancy over Entire Score Range for Significantly Different Comparisons of Blacks (B) or Spanish-surnamed (S) vs Whites*

Test-Pairs	Grade 4				Grade 5				Grade 6			
	N	Vocabulary	Comprehension	Total Reading	N	Vocabulary	Comprehension	Total Reading	N	Vocabulary	Comprehension	Total Reading
CAT - MAT									303 B		1.63 B	
CAT - STEP II					55 S	-1.73 S						
STEP II - CAT					55 S	2.30 S						
CAT - SAT					256 S	-1.08 S			286 S		1.75 S	
SAT - CAT					256 S	1.87 S		2.22 S				
CTBS - ITBS	293 B	1.28 B		1.59 B	295 B		1.73 B					
					98 S		3.20 S	3.73 S				
ITBS - CTBS	293 B	-1.39 B			295 B		-1.33 B					
					98 S		-2.18 S	-2.99 S				
CTBS - SAT	256 B			1.76 B								
SAT - CTBS	144 S	1.37 S			238 B			-2.10 B				
ITBS - CAT	85 B			2.91 B	90 B		1.83 B					
ITBS - MAT	117 S	2.48 S							205 B		-1.27 B	
MAT - SRA	457 B		-1.23 B	-1.67 B					539 B	-0.70 B		-1.23 B
SRA - MAT	457 B	2.40 B	1.82 B	3.48 B					539 B	1.15 B		1.71 B
MAT - SAT					107 S		2.39 S	3.77 S				
SAT - MAT	124 B	2.54 B		3.06 B	107 S		-2.06 S	-3.84 S				
STEP II - SAT					319 B	-1.67 B						
SAT - STEP II					319 B	1.80 B			68 S	1.48 S		
SRA - SAT	315 B		1.79 B									

*Data are entered for the group with the larger absolute discrepancy only.